**Objectives**

**Linking 16S sequences from Shotgun Libraries to Bins & Amplicon Profiles**

The 16S amplicon profiles obtained in a previous study were all trimmed to 102 bp. The cleaned reads range in length from 20-150 bp, but the majority are >130 bp. Bowtie2 was used with modified parameters to search reads for exact sequence matches to amplicon sequences. The non-default flags used include ` --score-min 'C,0,-1'`, `--all`, and `--no-unal` which together ensure that only when the entire query sequence is found in the reference, an alignment is recorded. The QC’d fastq files were filtered to remove all sequences <102 bp and converted to FASTA with seq-tk. The resultant SAM files were parsed by a custom script. During parsing it was recognized that two pairs of amplicons ('seq1799'/'seq50099' and 'seq83'/'seq86424') were aligning to the same reads, offset by a single nucleotide. Both alignments were added regardless.

The tags collected in the reverse and forward shotgun metagenomic were combined as the average of their relative abundances. This approach was also used for combining replicate samples collected in the amplicon dataset. The Pearson correlation between combined composition vectors was usually > 0.95 with the lowest few showing >0.74.

17 out of 712 amplicons were detected in the shotgun metagenomic library but not in the corresponding amplicon database. Most of these were detected at very low abundances, however seq86424 was an exception that comprised 24% and 20% of the 16S tags collected at 13 & 15 meters.

**Metagenomic Binning Procedure**

The requisite numerical information recommended by the Genomic Standards Consortium for metagenome-assembled genome (1) is shown in Table X. The following approaches apply to all bins derived. Taxonomic classification was performed using the taxator-tk algorithm (). Reassembly of bins and initial co-assembly was performed using SPAdes 3.11(). The initial assembly was completed in metagenomic mode and reassembly was done for each bin in `careful` mode using the first pass contigs as `untrusted contigs` .

Contigs were binned according to their coverage & tetramer frequencies. A set of consensus bins were derived from the bins produced by the maxbin2, metabat2, and concoct algorithms. Completeness and contamination assessment were performed using the lineage workflow in CheckM. The bioinformatics pipeline from QC, to assembly, to binning, to refinement, to reassembly, and taxonomic classification was done within the metaWRAP software. Prokka () was used to facilitate gene calling and preliminary annotations. Within prokka, Prodigal, barrnap, and Aragorn v1.2 were used to call ORFs, rRNA & tRNA respectively (,,). All protein sequences were generated using Translation Table 11.

Metabolic Model Processes

, but additional protein annotations were conducted using `metabolic-hmms` collection provided by the Banfield lab (), the ` dbCAN-fam-HMMs.v6` collection provided by BioEnergy Science Center of the DOE (), and the KEGG BlastKOALA and GhostKOALA annotation web service ().

1. Robert M Bowers, Nikos C Kyrpides, Ramunas Stepanauskas, Miranda Harmon-Smith, Devin Doud, T B K Reddy, Frederik Schulz, Jessica Jarett, Adam R Rivers, Emiley A Eloe-Fadrosh, Susannah G Tringe, Natalia N Ivanova, Alex Copeland, Alicia Clum, Eric D Becraft, Rex R Malmstrom, Bruce Birren, Mircea Podar, Peer Bork, George M Weinstock, George M Garrity, Jeremy A Dodsworth, Shibu Yooseph, Granger Sutton, Frank O Glöckner, Jack A Gilbert, William C Nelson, Steven J Hallam, Sean P Jungbluth, Thijs J G Ettema, Scott Tighe, Konstantinos T Konstantinidis, Wen-Tso Liu, Brett J Baker, Thomas Rattei, Jonathan A Eisen, Brian Hedlund, Katherine D McMahon, Noah Fierer, Rob Knight, Rob Finn, Guy Cochrane, Ilene Karsch-Mizrachi, Gene W Tyson, Christian Rinke, The Genome Standards Consortium, Alla Lapidus, Folker Meyer, Pelin Yilmaz, Donovan H Parks, A Murat Eren, Lynn Schriml, Jillian F Banfield, Philip Hugenholtz & Tanja Woyke. (2017). Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nature Biotechnology 35(725–731) doi:10.1038/nbt.3893
2. Kanehisa, M., Sato, Y., and Morishima, K. (2016) BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J. Mol. Biol. 428, 726-731 (http://www.kegg.jp/ghostkoala/)